DOCUMENT RESUME

ED 215 894 SE 037 098

AUTHOR O'Brien, Francis J., Jr.
TITLE A Proof That t2 and F are Identical: The General Case.
PUB DATE 24 Apr 82
NOTE 20p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Mathematics; Equations (Mathematics); Higher Education; Instructional Materials; *Mathematical Applications; Mathematical Concepts; *Mathematical Formulas; Mathematics; *Proof (Mathematics); *Research Tools; *Statistics; Supplementary Reading Materials

ABSTRACT
This document proves that the F statistic can be obtained by squaring t-test values, or that equivalent t-test values may be obtained by extracting the positive square roots of F values. Proof to varying degrees of completeness and accessibility has been given by other scholars, but generally these prior statements, particularly those available to students of education or psychology, focus on the special case, when sample sizes are equal. No source could be found that provided a complete, detailed proof of the general case that was understandable to students of applied statistics. This document seeks to give a clear step-by-step proof, with a numerical example worked out, and a plan is provided for proving the special case. It is felt the reader should be able to follow the proof of the general case, and should therefore have little difficulty in translating the acquired knowledge into proving the special case. (MP)

A Proof that t^2 and F are Identical: the General Case

Francis J. O'Brien, Jr., Ph.D.

It is well known that a researcher who has collected data from two independent groups may perform either a t-test for independent samples or a one-way analysis of variance for two groups. This is because knowledge of results from one type of computation can be transformed into an equivalent result for the other type of computation. For example, if a t-test for independent samples is calculated, the equivalent F statistic can be obtained by squaring the t-test value. Analogously, if the researcher has available the F statistic obtained from a one-way analysis of variance for two groups, the equivalent t-test value may be obtained by extracting the positive square root of the F value. That is, $t^2 = F$ or $t = (F)^{\frac{1}{2}}$.

The proof that $t^2 = F$ has been given to varying degrees of completeness and accessibility to students by other scholars in professional journals (See Rucci and Tweney, 1980 for citations). Statistics textbooks commonally available to students of education or psychology occasionally provide hints for proving the special case of the relationship (when sample sizes are equal). (See, for example, Glass and Stanley, 1970).

The motivation for presenting the proof is twofold. First, many prior statements of the proof for the general case (of unequal sample sizes) are either abbreviated, mathematically inaccessible or incomplete for understanding this important relationship. A search of the literature did not reveal a source that provided a complete, detailed proof of the general case that was understandable to students of applied statistics. Second, a full step-by-step proof for the general case will give readers a sense that statistics is not all just a babel of "Greek arithmetic". As a former instructor of graduate level applied statistics, I know that many students can follow well-articulated proofs and desire to see them worked out.

In this paper three tasks will be accomplished. First, a clear step-by-step proof that $t^2 = F$ in the general case will be provided. Second, a numerical example will be worked out. Third, a plan will be provided for proving the special case. It is felt that the reader should be able to follow the proof of the general case, and therefore, should have little difficulty in translating the acquired knowledge into proving the special case.

Proof that $t^2 = F$: the General Case

First let us lay out a table of symbolic values in order to introduce a familiar notation and the variables used in the proof. This is done in Table 1.

The plan for the proof is important for understanding the strategy involved in attacking a statistical proof. The steps of the plan that will be used here are given below:

1. state the form of the t-test statistic using the notation of Table 1.
2. square the t in step 1.
3. state the form of the F statistic using the notation of Table 1.
4. simplify algebraically the F statistic in step 3.
5. observe that the simplified F of step 4 is equal to the squared t of step 2.

Table 1

Table Layout for Two Independent Groups

	Group 1	Group 2	Total
	X_{11}	X_{12}	
	X_{21}	X_{22}	
	X_{31}	X_{32}	
	X_{41}	X_{42}	
	.	.	
	.	.	
	.	.	
	X_{i1}	X_{i2}	
	.	.	
	.	.	
	.	.	
	X_{n1}	X_{n2}	
Sample Size	$n._1$	$n._2$	$n..$
Sample Mean	$\bar{X}._1$	$\bar{X}._2$	$\bar{X}..$
Sample Variance	$s^2._1$	$s^2._2$	$s^2..$

Notes for Table 1

1. sample sizes are assumed unequal. That is,

   $n._1 \neq n._2$

2. the total sample size is

   $n.. = n._1 + n._2$

3. the grand mean ($\bar{x}..$) is a weighted mean since sample sizes are unequal. That is,

$$\bar{X}.. = \frac{n._1\bar{X}._1 + n._2\bar{X}._2}{n._1 + n._2}$$

4. $s^2..$ is not needed for the proof. It is included only for completeness.

This is the expression we will use as a basis of comparison with the simplified F statistic to be obtained in step 4. Note that the above squared t value is referred to simply as t^2.

Step 3. the F statistic

It will be recalled that the F statistic is the ratio of two independent sums of squares: a between[1] sums of squares (SS_b) and a within sums of squares (SS_w). Also, each sums of squares is divided by an appropriate degrees of freedom term: df_b (between sums of squares degrees of freedom) and df_w (within sums of squares degrees of freedom). The general expression of the F statistic (for any number of groups) is:

$$F_{J-1, n..-J} = \frac{SS_b/df_b}{SS_w/df_w}$$

The general form of df_b is "the number of groups minus one" (i.e., $df_b = J-1$, where J is the number of groups). For two groups, $J = 2$, and, so, df_b for two groups is $df_b = 2 - 1 = 1$. The general form of df_w is "the total sample size minus the number of groups (i.e., $df_w = n.. - J$). Since $n.. = n._1 + n._2$, we can write $df_w = n._1 + n._2 - J$. Hence, for J=2 groups, we can write $df_w = n._1 + n._2 - 2$.

---

[1] Grammarians will point out that the preposition "between" refers to the relationship of two entities while "among" refers to more than two. However, since the reference here is ultimately to two groups, "between' will be used instead of the correct "among": We accept the righteous indignation of the grammarian.

This is the expression we will use as a basis of comparison with the simplified F statistic to be obtained in step 4. Note that the above squared t value is referred to simply as t^2.

Step 3. the F statistic

It will be recalled that the F statistic is the ratio of two independent sums of squares: a between[1] sums of squares (SS_b) and a within sums of squares (SS_w). Also, each sums of squares is divided by an appropriate degrees of freedom term: df_b (between sums of squares degrees of freedom) and df_w (within sums of squares degrees of freedom). The general expression of the F statistic (for any number of groups) is:

$$F_{J-1,n..-J} = \frac{SS_{b}/df_b}{SS_{w}/df_w}$$

The general form of df_b is "the number of groups minus one" (i.e., $df_b = J-1$, where J is the number of groups). For two groups, $J = 2$, and, so, df_b for two groups is $df_b = 2 - 1 = 1$. The general form of df_w is "the total sample size minus the number of groups (i.e., $df_w = n.. - J$). Since $n.. = n._1 + n._2$, we can write $df_w = n._1 + n._2 - J$. Hence, for J=2 groups, we can write $df_w = n._1 + n._2 - 2$.

---

[1] Grammarians will point out that the preposition "between" refers to the relationship of two entities while "among" refers to more than two. However, since the reference here is ultimately to two groups, "between" will be used instead of the correct "among". We accept the righteous indignation of the grammarian.

Thus, using the notation of Table 1, we can write the F statistic for two groups as follows:

$$F_{1,n._1+n._2-2} = \frac{ss_{b/1}}{ss_{w/(n._1+n._2-2)}}$$

$$= \frac{ss_b}{\dfrac{ss_w}{(n._1+n._2-2)}}$$

In order to facilitate the proof, we will write F in the familiar terms of means and variances. That is:

$$F_{1,n._1+n._2-2} = \frac{n._1(\bar{X}._1-\bar{X}..)^2 + n._2(\bar{X}._2-\bar{X}..)^2}{\dfrac{(n._1-1)s._1^2 + (n._2-1)s._2^2}{n._1+n._2-2}}$$

$$= F$$

We will refer to this expression as simply F.

It is instructive to now compare the value of t^2 and F. For the readers convenience, we will restate t^2 and F so that they may be compared and referred to later. This is done in Table 2.

Table 2

Restatement of t^2 and F

t^2	F
$\dfrac{(\bar{X}_{.1}-\bar{X}_{.2})^2}{\dfrac{(n_{.1}-1)s_{.1}^2+(n_{.2}-1)s_{.2}^2}{n_{.1}+n_{.2}-2}\left(\dfrac{n_{.1}+n_{.2}}{(n_{.1})(n_{.2})}\right)}$	$\dfrac{n_{.1}(\bar{X}_{.1}-\bar{X}_{..})^2+n_{.2}(\bar{X}_{.2}-\bar{X}_{..})^2}{\dfrac{(n_{.1}-1)s_{.1}^2+(n_{.2}-1)s_{.2}^2}{n_{.1}+n_{.2}-2}}$

Step 4. Simplify F.

This is the next to the last step in the proof. A full step-by-step proof requires several algebraic steps. We will first simplify the numerator of F (SS_b). Notes pertaining to the algebraic manipulations are provided in the margin for the readers convenience. See the following page.

We start by looking at the numerator of F (SS_b).

$$SS_b = n_{\cdot 1}(\bar{X}_{\cdot 1}-\bar{X}_{\cdot\cdot})^2 + n_{\cdot 2}(\bar{X}_{\cdot 2}-\bar{X}_{\cdot\cdot})^2$$

$$SS_b = n_{\cdot 1}\left[\bar{X}_{\cdot 1} - \left(\frac{n_{\cdot 1}\bar{X}_{\cdot 1} + n_{\cdot 2}\bar{X}_{\cdot 2}}{n_{\cdot 1} + n_{\cdot 2}}\right)\right]^2 +$$

$$n_{\cdot 2}\left[\bar{X}_{\cdot 2} - \left(\frac{n_{\cdot 1}\bar{X}_{\cdot 1} + n_{\cdot 2}\bar{X}_{\cdot 2}}{n_{\cdot 1} + n_{\cdot 2}}\right)\right]^2$$

$$SS_b = n_{\cdot 1}\left[\frac{(n_{\cdot 1}+n_{\cdot 2})\bar{X}_{\cdot 1} - (n_{\cdot 1}\bar{X}_{\cdot 1} + n_{\cdot 2}\bar{X}_{\cdot 2})}{n_{\cdot 1} + n_{\cdot 2}}\right]^2 +$$

$$n_{\cdot 2}\left[\frac{(n_{\cdot 1}+n_{\cdot 2})\bar{X}_{\cdot 2} - (n_{\cdot 1}\bar{X}_{\cdot 1} + n_{\cdot 2}\bar{X}_{\cdot 2})}{n_{\cdot 1} + n_{\cdot 2}}\right]^2$$

Notes

Since,

$$\bar{X}_{\cdot\cdot} = \frac{n_{\cdot 1}\bar{X}_{\cdot 1} + n_{\cdot 2}\bar{X}_{\cdot 2}}{n_{\cdot 1} + n_{\cdot 2}},$$

we can substitute for $\bar{X}_{\cdot\cdot}$.

Finding common denominators for each bracketed term,

Removing inside parentheses, multiplying inside the group means and subtracting,

$$SS_b = n._1\left[\frac{n._1\bar{X}._1 + n._2\bar{X}._1 - n._1\bar{X}._1 - n._2\bar{X}._2}{n._1 + n._2}\right]^2 +$$

$$n._2\left[\frac{n._1\bar{X}._2 + n._2\bar{X}._2 - n._1\bar{X}._1 - n._2\bar{X}._2}{n._1 + n._2}\right]^2$$

Cancelling like terms,

$$SS_b = n._1\left[\frac{n._2\bar{X}._1 - n._2\bar{X}._2}{n._1 + n._2}\right]^2 +$$

$$n._2\left[\frac{n._1\bar{X}._2 - n._1\bar{X}._1}{n._1 + n._2}\right]^2$$

Factoring like sample size terms,

$$SS_b = n._1\left[\frac{n._2(\bar{X}._1 - \bar{X}._2)}{n._1 + n._2}\right]^2$$

$$n._2\left[\frac{n._1(\bar{X}._2 - \bar{X}._1)}{n._1 + n._2}\right]^2$$

Squaring each term separately inside the brackets,

$$SS_b = n_{\cdot 1}\left[\frac{(n_{\cdot 2})^2\ (\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2}{(n_{\cdot 1}+n_{\cdot 2})^2}\right] + n_{\cdot 2}\left[\frac{(n_{\cdot 1})^2\ (\bar{X}_{\cdot 2}-\bar{X}_{\cdot 1})^2}{(n_{\cdot 1}+n_{\cdot 2})^2}\right]$$

Rearranging terms by bringing all sample size terms outside brackets,

$$SS_b = \frac{(n_{\cdot 1})(n_{\cdot 2})^2\ (\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2}{(n_{\cdot 1}+n_{\cdot 2})^2} + \frac{(n_{\cdot 2})(n_{\cdot 1})^2\ (\bar{X}_{\cdot 2}-\bar{X}_{\cdot 1})^2}{(n_{\cdot 1}+n_{\cdot 2})^2}$$

Factoring out like sample size terms in numerator and denominator,

$$SS_b = \frac{(n_{\cdot 1})(n_{\cdot 2})}{(n_{\cdot 1}+n_{\cdot 2})^2}\left[n_{\cdot 2}(\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2 + n_{\cdot 1}(\bar{X}_{\cdot 2}-\bar{X}_{\cdot 1})^2\right]$$

Note that $(\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2$ is the same as $(\bar{X}_{\cdot 2}-\bar{X}_{\cdot 1})^2$ because two squared differences of the same terms result in the same quantity. Hence, we can factor within the brackets to obtain,

$$SS_b = \frac{(n_{\cdot 1})(n_{\cdot 2})}{(n_{\cdot 1}+n_{\cdot 2})^2}\left[(n_{\cdot 1}+n_{\cdot 2})\ (\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2\right]$$

Simplifying $(n_{\cdot 1}+n_{\cdot 2})$ and $(n_{\cdot 1}+n_{\cdot 2})^2$, we obtain,

$$SS_b = \frac{(n_{\cdot 1})(n_{\cdot 2})}{n_{\cdot 1}+n_{\cdot 2}}\ (\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2$$

This is just about it! Now substitute the value of SS_b just obtained into the F statistic, and obtain,

$$F = \frac{\dfrac{(n_{\cdot 1})(n_{\cdot 2})}{n_{\cdot 1}+n_{\cdot 2}} (\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2}{\dfrac{(n_{\cdot 1}-1)s^2_{\cdot 1} + (n_{\cdot 2}-1)s^2_{\cdot 2}}{n_{\cdot 1} + n_{\cdot 2} - 2}}$$

If we divide numerator and denominator by

$$\frac{(n_{\cdot 1})(n_{\cdot 2})}{n_{\cdot 1}+n_{\cdot 2}}$$ we obtain,

$$F = \frac{(\bar{X}_{\cdot 1}-\bar{X}_{\cdot 2})^2}{\dfrac{(n_{\cdot 1}-1)s^2_{\cdot 1} + (n_{\cdot 2}-1)s^2_{\cdot 2}}{n_{\cdot 1} + n_{\cdot 2} - 2}\left(\dfrac{n_{\cdot 1} + n_{\cdot 2}}{(n_{\cdot 1})(n_{\cdot 2})}\right)}$$

END OF PROOF.

Step 5. observe that $t^2 = F$.

If the value of F above is compared with value of t^2 it will be seen that they are in the same form, signifying that they are equal. Refer to Table 2 for this comparison.

This completes the entire proof in accordance with the five step plan. We now turn to a numerical example for two independent groups of unequal sample size.

Numerical Example

The analytic proof using algebraic rules for the general case was given in great detail. A numerical example should provide additional insight.

The data and descriptive statistics are provided in Table 3 which is modeled on Table 1. Note that the data were chosen for illustrative purposes only.

Table 3

Data for Numerical Example

	Group 1	Group 2	Total
	10	50	
	40	70	
	20	100	
	50	40	
	75		
	90		
Sample Size	6	4	10
Sample Mean	47.5	65.0	54.5
Sample Variance	957.5	700.0	NOT NEEDED

Refer to Table 2 for the t-test formula. Using the formula there, the t statistic value is:

$$t = \frac{47.5 - 65.0}{\sqrt{\frac{(6-1)957.5 + (4-1)700.0}{6 + 4 - 2}\left(\frac{10}{(6)(4)}\right)}}$$

$$t = \frac{-17.5}{\sqrt{\frac{6887.5}{8}\left(\frac{10}{24}\right)}}$$

$$t = -.9240$$

If we square this value, we obtain,

$$t^2 = (-.9240)^2$$

$$t^2 = \underline{\underline{.8537}} \qquad (t^2 = .8537)$$

We place the computed value in the margin for easy reference.

Now, compute the F statistic using the formula provided in Table 2.

$$F = \frac{6(47.5-54.5)^2 + 4(65.0-54.5)^2}{\frac{(6-1)957.5 + (4-1)700.0}{6 + 4 - 2}}$$

$$F = \frac{6(49) + 4(110.25)}{\frac{6887,5}{8}}$$

$$F = \underline{\underline{.8537}} \qquad (F = .8537)$$

Thus, $t^2 = .8537$ and $F = .8537$, or $t^2 = F$.

Proof of the Special Case

If the sample sizes are equal in a research design of two independent groups, the proof that $t^2 = F$ is somewhat easier to derive. Rather than perform the necessary algebraic manipulations, it may prove instructive for the reader to actually derive it himself or herself. Working through the analytic proof for the special case will solidify an understanding of the proof for the general case.

Some hints will be provided for the reader in proving the special case. They are summarized as follows.

1. $n_{\cdot 1} = n_{\cdot 2} = n$
2. $\bar{X}_{\cdot\cdot} = \dfrac{\bar{X}_{\cdot 1} + \bar{X}_{\cdot 2}}{2}$

That is, since sample sizes are equal, one symbol for sample size may be used; it could be called n. Also, since sample sizes are equal, the grand mean ($\bar{X}_{\cdot\cdot}$) is simply the average of the means of each group. This value for $\bar{X}_{\cdot\cdot}$ should be substituted in the proof.

By making these two changes and by following the five step outline used for the general case proof, the reader should be able to derive the proof for the special case. One may also wish to "make up" an easy to work with numerical data set to check on the process.

Note

For students or researchers who enjoy proofs in applied statistics, the following two references may be useful.

Edwards, Allen L. Expected Values of Discrete Random Variables and Elementary Statistics. New York: Wiley, 1964.

Guilford, J. P. and Fruchter, Benjamin. Fundamental Statistics in Psychology and Education, 4th and 5th editions. New York: McGraw-Hill, 1973.

References

Glass, Gene V and Stanley, Julian C. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Rucci, Anthony J. and Tweney, Ryan D. Analysis of variance and the "second discipline" of scientific psychology: a historical account. Psychological Bulletin, 1980, Vol. 87, No. 1, 166-184.